DOCUMENT RESUME

ED 093 916 TM 003 733

AUTHOR Breland, Hunter M.
TITLE An Investigation of Cross-Cultural Stability in Mental Test Items.
PUB DATE [Apr 74]
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Complexity Level; Cross Cultural Studies; *Item Analysis; *Statistical Analysis; *Test Bias

ABSTRACT

Examples of cross-cultural stability or instability of mental test items are illustrated. A statistical procedure involving the cross-plotting of item difficulties for two different groups and generating a line of mutual regression through the resulting scatter of points is described. D-values, representing the perpendicular distance, in delta units, from the line of mutual regression, for all items in a given subtest (in this case, mathematics) are presented in a computer generated table for the following groups studied: American Indian, Afro-American, Mexican-American, Puerto Rican, other Latin-American, Oriental-American, White Eastern, White Southern, and White Western. (Author/BC)

TM 003 733 — ED 093916

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AN INVESTIGATION OF CROSS-CULTURAL STABILITY IN MENTAL TEST ITEMS

Hunter M. Breland
Educational Testing Service
Princeton, New Jersey

Paper presented at the Annual Meeting
of the American Educational Research Association,
Chicago, Illinois, April, 1974.

AN INVESTIGATION OF CROSS-CULTURAL STABILITY IN MENTAL TEST ITEMS[1]

Hunter M. Breland

Educational Testing Service

For purposes of illustrating what I intend by cross-cultural stability (or instability) in a mental test item, I want to present some examples. Imagine that you are back in the first grade and the teacher is administering a test to determine what you understand about words. A booklet is given to you and each page has two columns of pictures. The teacher instructs you to mark with a big X either the picture in the left hand column or the picture in the right hand column as she pronounces a word representing one of the pictures. One of the picture-pairs has on the left a common device for use with baby diapers and on the right a common writing implement, as shown at the top in Figure 1. Now imagine that I am the teacher. I will pronounce the word and you are to put a big X across the picture that the word represents. Are you ready? (I pronounce the word). Most of you would mark the picture on the left because, in your socio-cultural group, the word representing the picture on the right is pronounced differently than the word representing the picture on the left. In the southern part of the United States, however, the words are pronounced the same. The two pictures are distinguished by preceding the "common" word by another word. That is,

---

1. This paper is based upon results obtained in a study prepared for The National Institute of Education, Breland and Stocking (1974). (Project No. 3-0658, Grant No. NE-6-00-3-0116)

the picture on the left is described as a "safety pen" and the picture on the right as a "fountain pen." But in both cases the word "pen" is pronounced as "pin." You will have to admit that both have sharp points! What I have just illustrated is a very subtle socio-cultural difference in the way the same language is used.

All cross-cultural instabilities in mental test items, however, are not due to socio-linguistic phenomena. Thorndike (1971) presented a hypothetical item about baking (the second item in Figure 1). Most housewives would know that the correct answer is (c). But most men, boys, and young girls not yet introduced to cooking would not. Thus, whether one is likely to know the correct answer is a function of his group membership. As Thorndike pointed out, whether such an item in unfair to a given group depends upon what it is being used for. If used in a test to select cooks, then one might find that the cooks selected were mostly women (assuming the test used a considerable number of similar items).

The above two examples are rather obvious ones. Let me show you something similar, but not obvious at all. Consider the third item in Figure 1. This is a vocabulary item in which a stem word is presented and the task is to identify the option most closely the same in meaning. In the investigation I am about to report, 50% of groups of high school seniors identified as either North-Central Caucasians or Southern Caucasians responded (A), the correct answer. But in a group of Latin-Americans living in the United States, and among whom half reported that English was not the language spoken most often in their homes, 56% responded correctly. All groups being compared were from carefully selected national random samples. A close look at this item, from a cross-cultural perspective, revealed that

both the stem word, convalesce, and the correct option, recuperate, have cognates in the Spanish language—convalecer and recuperar.

I have presented three very different types of mental test items, all of which demonstrate some kind of cross-cultural instability. The first item was created from my own experiences as a transplanted Southerner living and working in a foreign culture (New Jersey). The second was obtained from a perceptive educational researcher. But the last item was discovered by means of a statistical procedure.

This procedure, from Angoff and Ford (1973) involves the cross-plotting of item difficulties for two different groups and generating a line of mutual regression through the resulting scatter of points. An example is presented in Figure 2, where a national random sample of Oriental-Americans is compared to a random sample of North-Central American Caucasians.

Deltas are obtained by the transformation

$$\Delta = 4Z + 13.$$

The dotted line in Figure 2 is drawn at 45° as a line of reference. The solid line is the line of mutual regression (least squares are based on perpendicular distances). The item-points (+'s) represent most of the items from a six-test battery used in the Base-Year Survey of the National Longitudinal Study[2] (Hilton & Rhett, 1973). The six tests were: Vocabulary, Picture-Number, Reading, Letter-Groups, Mathematics, and Mosaic Comparisons.

Since the line of mutual regression falls above the dotted reference line, the suggestion is that the battery as a whole was easier for Oriental-Americans than it was for North-Central American Caucasians. What is of primary interest in the present paper, however, are the specific items that

---

[2] Conducted by ETS for the U. S. Office of Education.

were either easier or more difficult for a given group than other items in the battery. Some of the item-points are some distance from the solid line and thus are cause for further inquiry. An indication of what characterizes the items that are relatively easier for Oriental-Americans is given by the computer-generated Table 1, which presents what are termed D-values for all items in a given subtest (in this case, Mathematics) and for all groups studied. The groups are identified in Table 1 as follows: AI-American Indian, AA-Afro-American, MA-Mexican-American, PR-Puerto Rican, OL-other Latin-American, OR-Oriental-American, WE-White Eastern, WS-White Southern, and WW-White Western. The White North-Central group is not shown in the Table because it is the group with which all others are compared.

The D-values represent the perpendicular distance, in delta units, from the line of mutual regression (the solid line in Figure 2). Negative D-values represent distances above the solid line and positive D-values distance below the solid line. Therefore, negative D-values indicate that an item is **relatively easier for the group plotted on the abscissa** (in the case of Figure 2 and the column headed, or, in Table 1, Oriental-Americans). The relative ease of the Mathematics test, as a whole, in contrast to the entire six-test battery, as a whole, is suggested by the Group Mean at the bottom of Table 1. The mean D-value of -.33 for Oriental-Americans would suggest that the Mathematics test was easier for this group than other tests in the battery. The individual D-values indicate, however, that all items on the Mathematics subtest were not easier for Oriental-Americans. Item 6 appears to have been substantially harder. Item 16 was one of the easiest for Oriental-Americans. Consider the contrast of these two items:

Item 6 (difficult for Oriental-Americans)		Item 16 (easy for Oriental-Americans)	
A	B	A	B
Degree rise in temperature from -6°F to +5°F	Degree rise in temperature from -5°F to +6°F	$\frac{1}{2} \div \frac{1}{3}$	$\frac{1}{2} \times \frac{1}{3}$

The task in this kind of item is to determine which of the following statements are true: (1) The quantity in Column A is greater, (2) The quantity in Column B is greater, (3) The two quantities are equal, or (4) The size relationship cannot be determined from the information given. For Item 6, even though 72% of the Oriental-Americans responded correctly with response (3), that the two quantities are equal, the D-values of Table 1 show that Item 6 was still relatively more difficult for them than other items in the six-test battery. All of the Caucasian groups either exceeded or approached 80% correct on Item 6.

In contrast, 78% of Oriental-Americans correctly chose response (1) for Item 16, that the quantity in Column A is greater than the quantity in Column B. But only about 60% of the Caucasian groups responded correctly, with the White Eastern group slightly better than the White Western group (63% to 58%). One thing that characterizes the difference in these two items is the degree of verbalization. Item 6, which was relatively difficult for Oriental-Americans, has a lot of verbalization; Item 16 has none at all. In the entire Mathematics subtest, 12 of the 25 items involved no verbalization: Items 1, 4, 5, 8, 10, 12, 13, 16,

17, 19, 21, and 25. Looking at Table 1, it is seen that ten of these 12 have large negative D-values for Oriental-Americans. Only Items 1 and 4 do not fit this pattern.

There is not time here to discuss a number of other interesting observations made using a procedure similar to that described. I will mention them briefly. The statistical procedures detected a set of items within the Reading subtest that appeared relatively easier for Blacks than other items in the battery. On investigation, these items all were revealed to be associated with a reading passage about Black television. For minorities in general (with the exception of Oriental-Americans), the contrast of Item 2 and Item 5 in Table 1 is revealing. Item 2 dealt with money, something that might be learned outside of school. The large negative D-values for all minorities in Table 1 suggest that Item 2 is relatively easier for them than most other items in the six-test battery. But Item 5 appears to be relatively more difficult for minorities. The large positive D-values for minorities in Table 1 contrast with negative D-values for all Caucasian groups and Oriental-Americans. What is so striking about this contrast is the simplicity of Item 5—it merely asks to indicate which of the quantities, $\sqrt{9}$ or 9, is greater. While such an item is simple, it represents a type of knowledge usually only obtained in school. That minorities should perform so poorly on this type of item suggests that the schools they attend are providing very poor instruction in mathematics. While 72% of American Indians got Item 2 correct (which is greater, 11 dimes or $1.11),

only 42% got Item 5 correct. On the next item, Item 6, 64% of American Indians responded correctly--indicating that the difference between Items 2 and 5 is not a speed factor. When it is considered that 25% of the 42% correct for American Indians on Item 5 could have been attained by guessing alone (and only 5 of the sample of 178 American Indians failed to respond to Item 5), one suspects that minority groups in this country are taught very little about mathematics in high school. About 75% of the Caucasian groups and 85% of Oriental-Americans responded correctly to Item 5.

In conclusion, what started as an investigation of mental tests has ended as an investigation of high school instruction in the United States. Although tests appear to have some problems, they also are useful in evaluating instruction.

References

Angoff, W. H., & Ford, S. F. Item-race interaction on a test of scholastic aptitude. Journal of Educational Measurement, 1973, 10(2), 95-105.

Breland, H. M., & Stocking, M. The cross-cultural stability of mental test items: An investigation of response patterns for ten socio-cultural groups. Final Report, National Institute of Education Project No. 3-0658 (Grant No. NE-G-00-3-0116), Project Report PR-74-2, Educational Testing Service, Princeton, N. J., February, 1974.

Hilton, T. L., & Rhett, H. The Base-Year Survey of the National Longitudinal Study of the High School Class of 1972. Final Report, Contract No. OEC-0-72-0903, Office of Education, National Center for Educational Statistics, U. S. Department of Health, Education and Welfare. Educational Testing Service, Princeton, N. J., June, 1973.

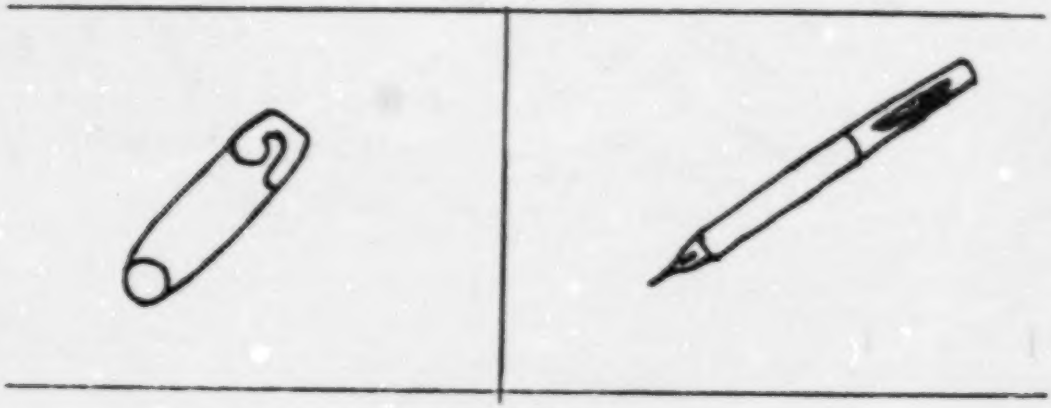
Thorndike, R. L. Concepts of culture-fairness. Journal of Educational Measurement, 1971, 8, 63-70.

Figure 1

Examples of Cross-Culturally Unstable

Mental Test Items

1.



2. The usual temperature for baking a cake is about:

   (A) 250° (B) 300° (C) 350° (D) 400°

3. Select the one word or phrase whose meaning is closest to that of the word in capital letters.

   CONVALESCE:

   (A) recuperate
   (B) bring together
   (C) condemn
   (D) appreciate
   (E) collect

Figure 2

Cross-plot of Deltas for Oriental

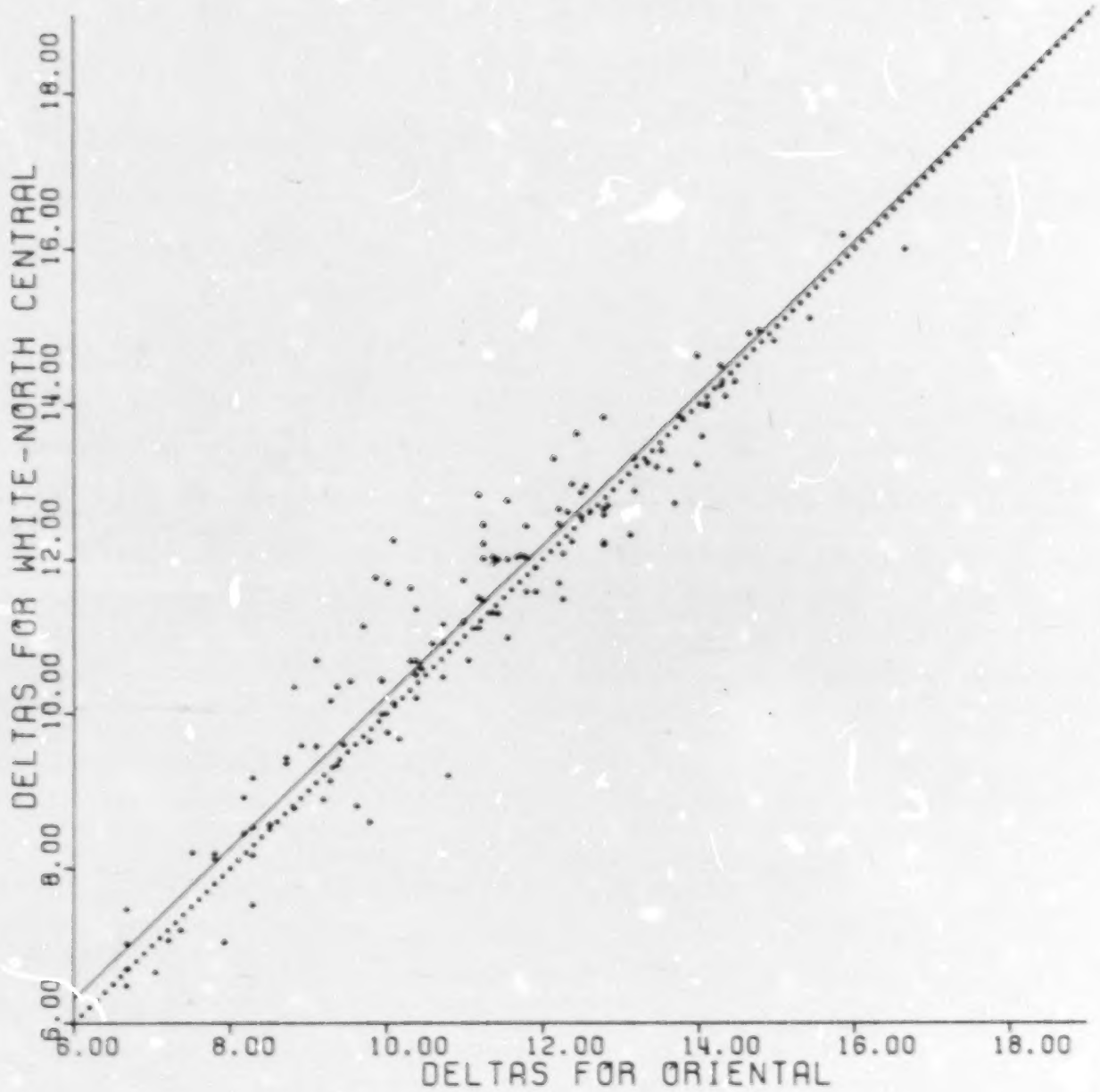


TABLE 1
ITEM D-VALUES BY GROUP
MATHEMATICS

ITEM	AI	AA	MA	PR	OL	OR	WE	WS	WW	ITEM MEAN	ITEM S.D.
I- 1	0.50	0.35	0.36	0.22	0.59	0.02	0.04	0.09	0.29	0.33	0.20
I- 2	-0.83	-1.04	-0.91	-0.90	-1.27	-0.46	-0.20	-0.18	-0.03	-0.65	0.42
I- 3	0.09	0.06	-0.05	0.23	-0.05	0.21	-0.18	-0.24	-0.06	-0.00	0.15
I- 4	0.25	0.05	-0.04	0.14	0.03	0.39	0.20	0.04	-0.20	0.10	0.17
I- 5	1.36	0.47	0.54	0.72	0.51	-0.91	-0.08	-0.20	-0.05	0.26	0.61
I- 6	0.53	0.54	0.92	0.46	0.98	1.29	0.31	0.35	0.38	0.64	0.32
I- 7	0.03	0.31	0.68	0.51	0.49	0.07	0.01	0.21	0.10	0.27	0.23
I- 8	0.22	0.28	0.34	-0.14	0.31	-0.79	-0.05	-0.07	0.00	0.01	0.33
I- 9	0.12	-0.14	-0.09	0.53	0.04	0.12	-0.10	-0.02	0.06	0.06	0.19
I-10	0.05	0.17	0.14	0.25	0.26	-1.04	-0.00	-0.15	-0.01	-0.04	0.38
I-11	-0.14	0.35	0.22	0.26	-0.07	0.50	0.06	0.03	0.10	0.14	0.19
I-12	0.79	-0.28	-0.08	-0.14	0.21	-0.96	-0.09	-0.07	-0.16	-0.09	0.43
I-13	0.48	0.39	0.27	0.85	-0.03	-0.86	-0.10	0.05	0.10	0.13	0.45
I-14	0.03	0.66	0.17	0.21	-0.26	-0.06	0.07	0.16	-0.02	0.11	0.24
I-15	0.21	-0.21	0.49	-0.15	0.77	-0.27	0.11	0.09	0.16	0.13	0.32
I-16	0.84	0.34	0.01	0.02	0.01	-1.39	-0.18	-0.35	0.00	-0.08	0.56
I-17	0.29	-0.09	0.05	0.68	-0.20	-1.20	-0.09	-0.16	-0.02	-0.08	0.47
I-18	0.25	0.13	0.41	0.43	0.83	0.22	0.06	0.15	0.10	0.29	0.23
I-19	0.23	-0.10	-0.04	0.25	-0.70	-0.54	0.14	0.11	0.18	-0.05	0.33
I-20	0.75	0.78	0.85	1.02	0.84	0.29	0.07	0.09	0.15	0.54	0.36
I-21	0.72	0.91	0.41	0.64	-0.03	-1.04	-0.05	0.01	0.32	0.21	0.55
I-22	0.82	0.86	0.69	0.47	0.17	-0.73	0.08	0.18	0.07	0.29	0.47
I-23	0.22	0.17	0.34	0.15	0.48	-0.64	-0.04	-0.04	0.04	0.08	0.30
I-24	0.57	0.66	0.42	1.05	0.60	0.19	0.09	0.07	0.14	0.42	0.31
I-25	0.27	-0.11	-0.11	-0.17	0.11	-0.73	-0.06	0.04	0.30	-0.05	0.29
GROUP MEAN	0.35	0.22	0.26	0.30	0.18	-0.33	0.00	0.01	0.08		
S.D.	0.412	0.418	0.394	0.419	0.490	0.640	0.119	0.156	0.141		

***** D-VALUES NOT COMPUTED FOR MISSING ITEM-DELTAS